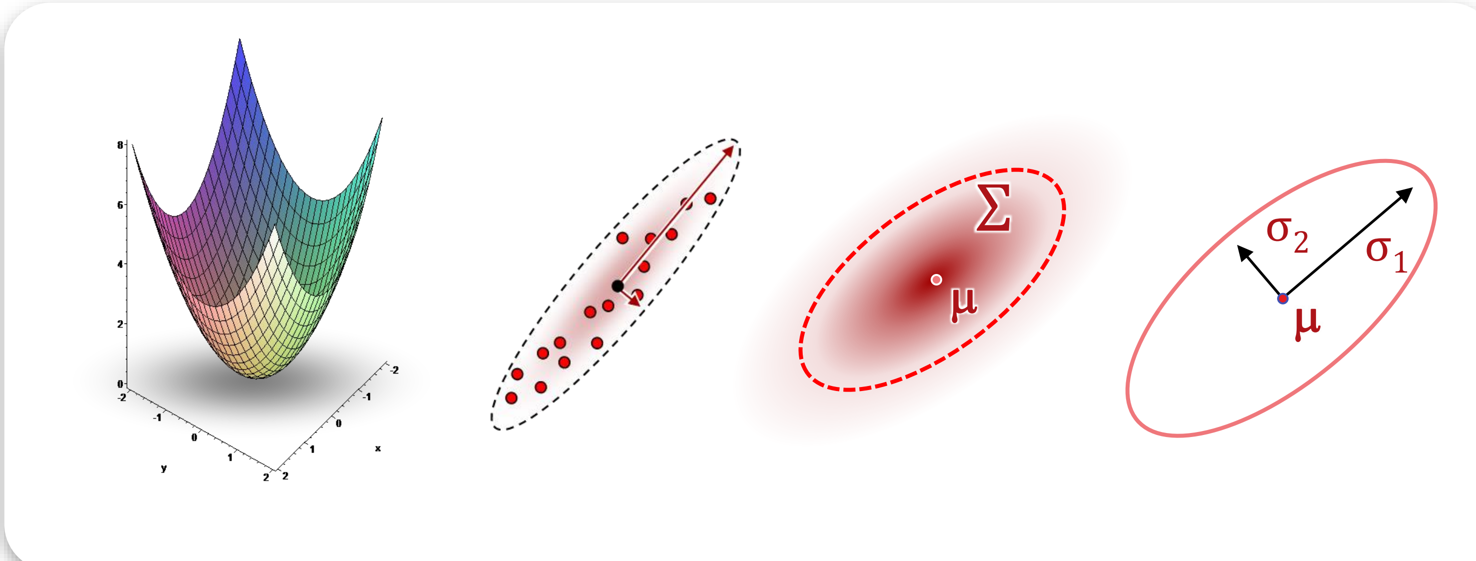
Modelling 1

SUMMER TERM 2020



LECTURE 16

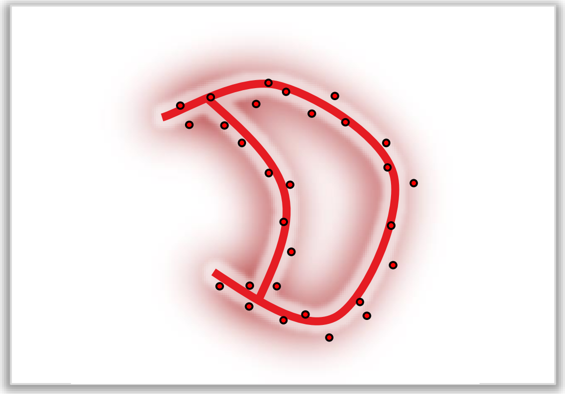
Total Least Squares & PCA

Michael Wand · Institut für Informatik · Michael.Wand@uni-mainz.de

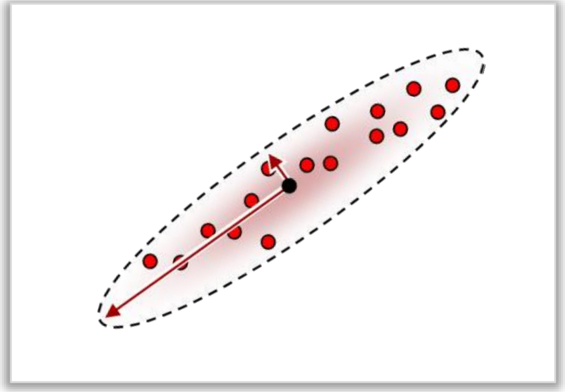
Three Different Stories!

PCA, told in 3 different ways…

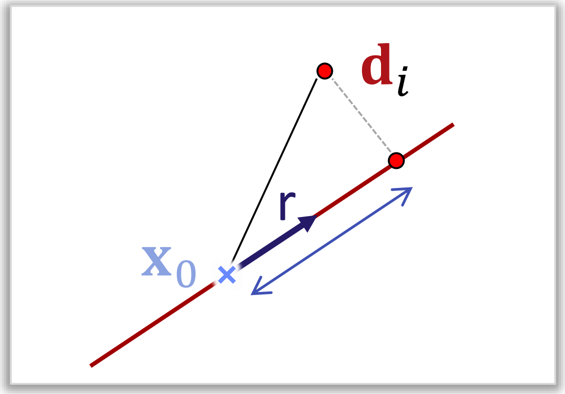
- **Total least-squares:**
  Euclidean error



- **Multivariate Gaussians:**
  Main Axes of Variance



- **Dimensionality Reduction:**
  Optimal projection to subspace

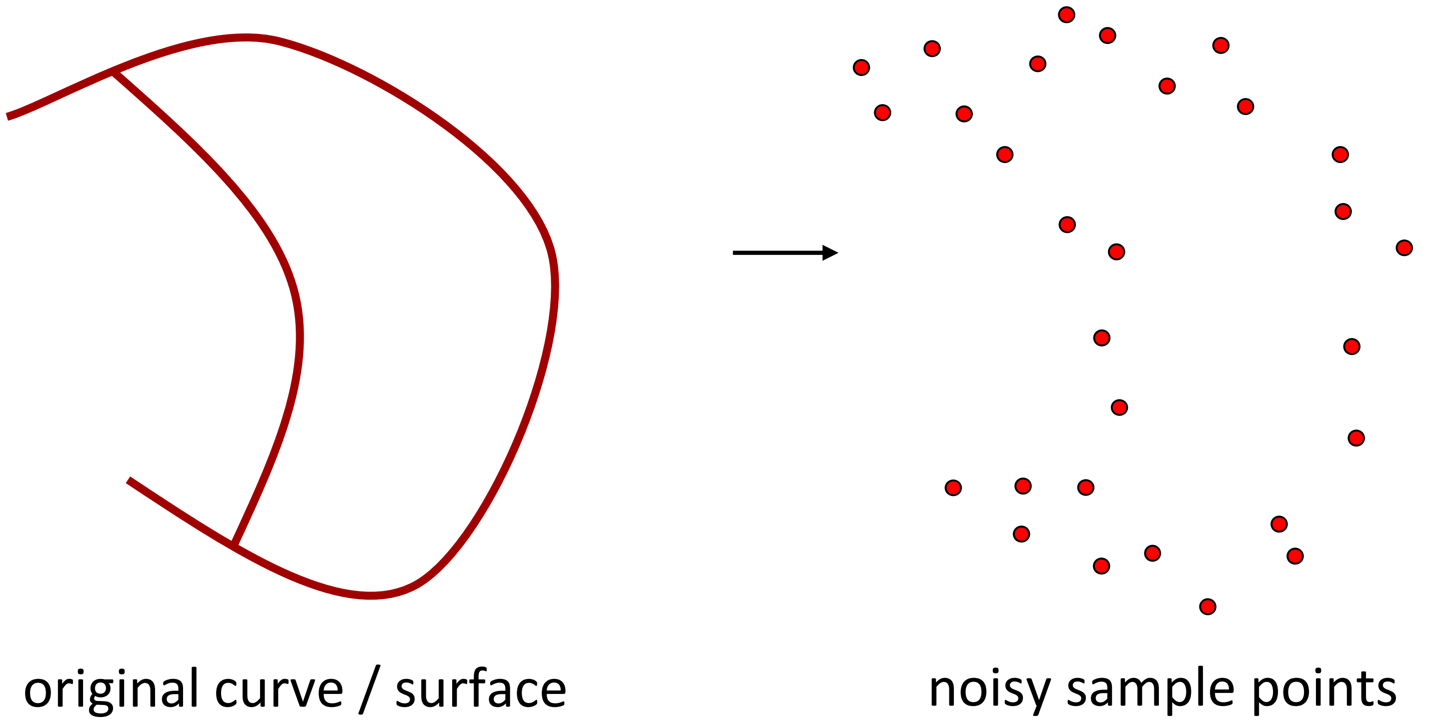


Story 1

Total Least Squares

Statistical Model

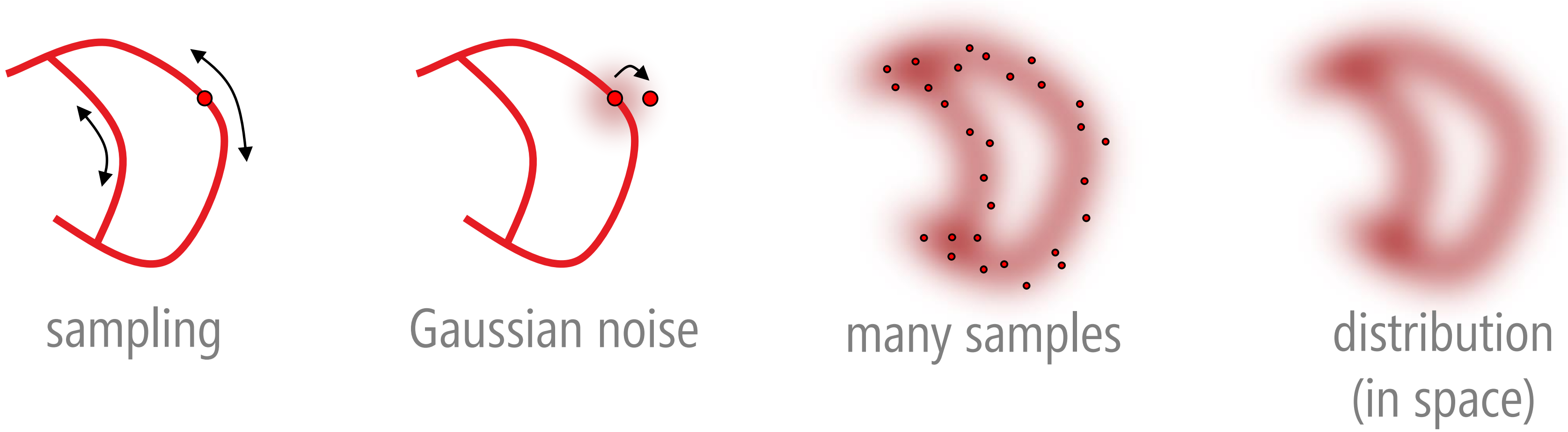
Generative Model:



Statistical Model

Generative Model:

1. Determine sample point (uniform)
2. Add noise (Gaussian)



Squared Distance Function

Result:

- Gaussian distribution convolved with object
- No analytical density

Approximation:

- 1D Gaussian $\rightarrow$ minimize squared residual
- This case $\rightarrow$ minimize squared distance function

General Total Least Squares

General Total Least Squares:

- Given:
  - Choice of objects S_λ, by parameters $\boldsymbol{\lambda} \in \mathbb{R}^k$
  - Set of n sample points $\mathbf{d}_i \in \mathbb{R}^d$ (Gaussian, iid, isotropic).
- Total least squares minimizes

$$\underset{\lambda \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n} \operatorname{dist}(S_\lambda, \mathbf{d}_i)^2$$

- In general: Non-linear optimization problem
  - Special cases can be solved exactly

Fitting Affine Subspaces

The following problem can be solved exactly:

- Best fitting line to a set of 2D, 3D points
- Best fitting plane to a set of 3D points
- ...
- In general
  - Affine subspace of $\mathbb{R}^d$
  - Dimension $d_{sub} \leq d$
  - Approximates a set of data points $\mathbf{d}_i \in \mathbb{R}^m$ best
    - in a sum-of-squared-distances sense

Solution: principle component analysis (PCA)

Start: 0-dim Subspaces

Optimal 0-dimensional affine subspace

- Given
  - Set $\mathbf{D}$ of n data points $\mathbf{d}_i \in \mathbb{R}^d$,
- Looking for:
  - Point $\mathbf{x}_0$ with minimum sum-of-squared distances (to all data points)
- Answer: just the mean (a.k.a. average):

$$\mathbf{x}_0^{(opt)} = m(\mathbf{D}) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i$$

- Proof: minimize $E(\mathbf{x}_0) = \sum_{i=1}^{n} \|\mathbf{x}_0 - \mathbf{d}_i\|^2$ (next slide...)

Proof

$$E(\mathbf{x}_0) = \sum_{i=1}^{n} \|\mathbf{x}_0 - \mathbf{d}_i\|^2$$

$$= \sum_{i=1}^{n} \|\mathbf{x}_0 - m(\mathbf{D}) + m(\mathbf{D}) - \mathbf{d}_i\|^2$$

$$= \sum_{i=1}^{n} \left(\mathbf{x}_0 - m(\mathbf{D})\right)^2 - 2\sum_{i=1}^{n} \langle \mathbf{x}_0 - m(\mathbf{D}), m(\mathbf{D}) - \mathbf{d}_i \rangle + \sum_{i=1}^{n} (m(\mathbf{D}) - \mathbf{d}_i)^2$$

$$= \sum_{i=1}^{n} \left(\mathbf{x}_0 - m(\mathbf{D})\right)^2 - 2\left\langle \mathbf{x}_0 - m(\mathbf{D}), \sum_{i=1}^{n} (m(\mathbf{D}) - \mathbf{d}_i) \right\rangle + \sum_{i=1}^{n} (m(\mathbf{D}) - \mathbf{d}_i)^2$$

$$= \sum_{i=1}^{n} \left(\mathbf{x}_0 - m(\mathbf{D})\right)^2 - 2\left\langle \mathbf{x}_0 - m(\mathbf{D}), nm(\mathbf{D}) - \sum_{i=1}^{n} \mathbf{d}_i \right\rangle + \sum_{i=1}^{n} (m(\mathbf{D}) - \mathbf{d}_i)^2$$

$$= \sum_{i=1}^{n} \left(\mathbf{x}_0 - m(\mathbf{D})\right)^2 - 2\langle \mathbf{x}_0 - m(\mathbf{D}), nm(\mathbf{D}) - nm(\mathbf{D}) \rangle + \sum_{i=1}^{n} (m(\mathbf{D}) - \mathbf{d}_i)^2$$

$$= \underbrace{\sum_{i=1}^{n} \left(\mathbf{x}_0 - m(\mathbf{D})\right)^2}_{\text{minimal for } \mathbf{x}_0 = m(\mathbf{D})} + \underbrace{\sum_{i=1}^{n} (m(\mathbf{D}) - \mathbf{d}_i)^2}_{\text{independent of } \mathbf{x}_0}$$

Sample mean:

$$m(\mathbf{D}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{d}_i$$

One Dimensional Subspaces...

First step

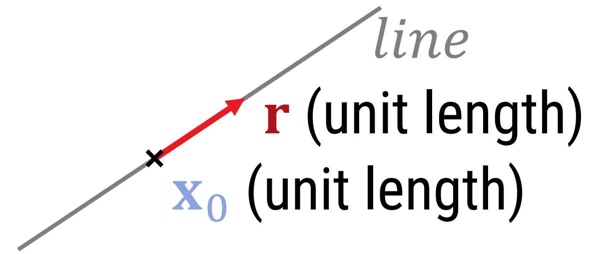
- Optimal (1D) line approximating $\{\mathbf{d}_1, \dots, \mathbf{d}_n\}$?
- Two questions:
  - Optimum base point?
    – Still the average (proof omitted)
  - Optimum direction?
    – We will look at that next...
- Parametric line equation:

$$\mathbf{x}(t) = \mathbf{x}_0 + t \cdot \mathbf{r} \qquad (\mathbf{x}_0 \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d, ||\mathbf{r}|| = 1)$$

Best Fitting Line

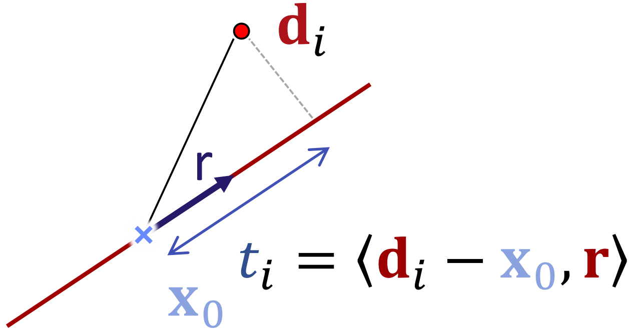
Parametric line equation:

$$\mathbf{x}(t) = \mathbf{x}_0 + t \cdot \mathbf{r} \quad \left(\mathbf{x}_0 \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d, \|\mathbf{r}\| = 1\right)$$



Best projection onto line

$$t_i = \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle$$



Objective Function

$$E(\mathbf{r}) = \sum_{i=1}^{n} \mathrm{dist}(line, \mathbf{d}_i)^2 = \sum_{i=1}^{n} \left([\mathbf{x}_0 + t_i \mathbf{r}] - \mathbf{d}_i\right)^2$$

Objective Function

Least-Squares Error: (sum of squared distances)

$$E(\mathbf{r}) = \sum_{i=1}^{n} ([\mathbf{x}_0 + t_i \mathbf{r}] - \mathbf{d}_i)^2$$

rebracket

$$= \sum_{i=1}^{n} (t_i \mathbf{r} - [\mathbf{d}_i - \mathbf{x}_0])^2$$

rewrite scalar product

$$= \sum_{i=1}^{n} (t_i \mathbf{r} - [\mathbf{d}_i - \mathbf{x}_0])^{\mathrm{T}} (t_i \mathbf{r} - [\mathbf{d}_i - \mathbf{x}_0])$$

expand, split sum

$$= \sum_{i=1}^{n} t_i^2 \mathbf{r}^{\mathrm{T}} \mathbf{r} - 2 \sum_{i=1}^{n} t_i \mathbf{r}^{\mathrm{T}} [\mathbf{d}_i - \mathbf{x}_0] + \sum_{i=1}^{n} [\mathbf{d}_i - \mathbf{x}_0]^2$$

Objective Function

Least-Squares Error: (sum of squared distances)

$$E(\mathbf{r}) = \sum_{i=1}^{n} t_i^2 \mathbf{r}^{\mathrm{T}}\mathbf{r} - 2\sum_{i=1}^{n} t_i \mathbf{r}^{\mathrm{T}}[\mathbf{d}_i - \mathbf{x}_0] + \sum_{i=1}^{n} [\mathbf{d}_i - \mathbf{x}_0]^2$$

= 1 (||r||=1)

now: insert

$$t_i = \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle$$

$$E(\mathbf{r}) = \sum_{i=1}^{n} \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle^2 - 2\sum_{i=1}^{n} \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle \mathbf{r}^{\mathrm{T}}[\mathbf{d}_i - \mathbf{x}_0] + \sum_{i=1}^{n} [\mathbf{x}_0 - \mathbf{d}_i]^2$$

$$= \sum_{i=1}^{n} \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle^2 - 2\sum_{i=1}^{n} \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle^2 + \sum_{i=1}^{n} [\mathbf{x}_0 - \mathbf{d}_i]^2$$

$$= -\sum_{i=1}^{n} \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle^2 + \sum_{i=1}^{n} [\mathbf{x}_0 - \mathbf{d}_i]^2$$

(const wrt. r)

Objective Function

We got:

$$E(\mathbf{r}) = \sum_{i=1}^{n} \langle \mathbf{d}_i - \mathbf{x}_0, \mathbf{r} \rangle^2 + \text{const.}$$

So we have to optimize:

$$E(\mathbf{r}) = -\mathbf{r}^{\mathrm{T}} \cdot \left[\sum_{i=1}^{n} (\mathbf{d}_i - \mathbf{x}_0)(\mathbf{d}_i - \mathbf{x}_0)^{\mathrm{T}} \right] \cdot \mathbf{r} + \text{const.}$$

$$\text{subject to: } \|\mathbf{r}\| = 1$$

Best Fitting Line

Result:

$$\sum_{i=1}^{n} \text{dist}(line, \mathbf{d}_i)^2 \sim -\mathbf{r}^{\mathrm{T}} \cdot \mathbf{S} \cdot \mathbf{r} + \text{const.}$$

$$\text{with} \quad \mathbf{S} := \sum_{i=1}^{n} (\mathbf{d}_i - \mathbf{x}_0)(\mathbf{d}_i - \mathbf{x}_0)^{\mathrm{T}}, \qquad \text{and s.t.} \|\mathbf{r}\| = 1$$

Eigenvalue Problem:

$$\mathbf{r}^{\mathrm{T}} \cdot \mathbf{S} \cdot \mathbf{r} \;\; \text{with} \;\; \|\mathbf{r}\| = 1 \quad \leftrightarrow \quad \frac{\mathbf{r}^{\mathrm{T}} \cdot \mathbf{S} \cdot \mathbf{r}}{\mathbf{r}^{\mathrm{T}} \cdot \mathbf{r}}$$

- Rayleigh quotient
  - Minimizing the energy: maximum quotient
  - Solution: eigenvector with *largest* eigenvalue

General Case

Fitting a d-dimensional affine subspace:

- $d = 1$: line
- $d = 2$: plane
- $d = 3$: 3D subspace
- ...

Optimal approximation

- Use d eigenvectors with the *largest eigenvalues*
- Yields (total) least-squares optimal subspace for approximating data $\{\mathbf{d}_1, \ldots, \mathbf{d}_n\}$

General Case

Principal Component Analysis (PCA)

- Compute average

$$\mathbf{x}_0 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i$$

- Compute “scatter matrix”

$$\mathbf{S}(\mathbf{d}_1, \ldots, \mathbf{d}_n) = \sum_{i=1}^{n} (\mathbf{d}_i - \mathbf{x}_0)(\mathbf{d}_i - \mathbf{x}_0)^{\mathrm{T}}$$

General Case

Principal Component Analysis (PCA)

- $(\lambda_1, \mathbf{v}_1), \dots, (\lambda_n, \mathbf{v}_n)$: sorted eigenvalue/vector pairs of $\mathbf{S}$
  - λ_1 is the largest
  - $\|\mathbf{v}_1\| = 1$
- Select subspace spanned by

$$\mathbf{x}_0 + \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_d\}, \qquad 0 \leq d \leq n$$

- Subspace-projection is optimal:
  - Yields optimal d-dim approximation among all possible affine subspaces (Wrt. squared distances)

Story 3: Linear Dimensionality Reduction

Remember the…

Gaussians

(Story 2…)

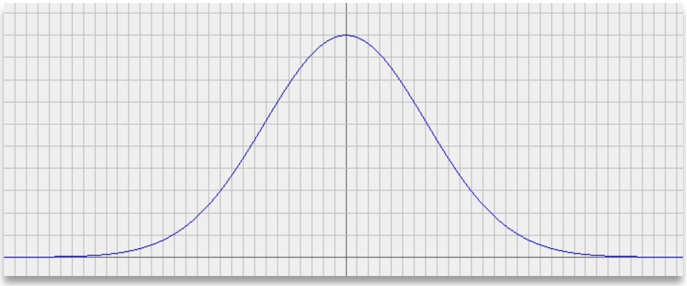
Gaussians

Gaussian Normal Distribution

- Two parameters: μ, σ
- Density:

$$\mathcal{N}_{\mu,\sigma}(x) \coloneqq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean: μ
- Variance: σ^2



Gaussian normal distribution

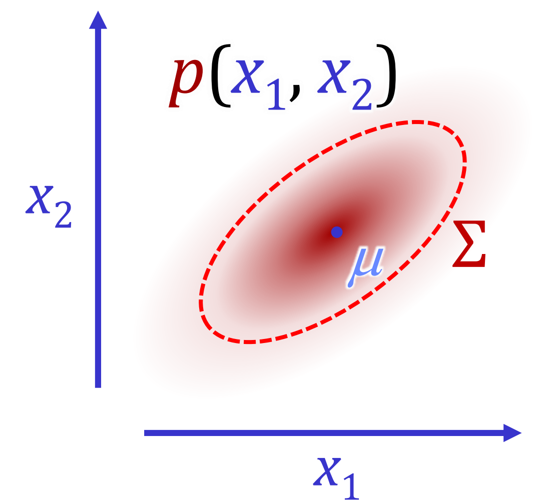
Multi-Variate Gaussians

Gaussian Normal Distribution in d Dimensions

- Two parameters: $\boldsymbol{\mu}$ (d-dim-vector), $\boldsymbol{\Sigma}$ ($d \times d$ matrix)
- Density:

$$\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) := \left(\frac{1}{(2\pi)^{-\frac{d}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}}} \right) e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

- Mean: $\boldsymbol{\mu}$
- Covariance Matrix: $\boldsymbol{\Sigma}$



Estimation from Data

Task

- Data (i.i.d.) $\mathbf{d}_1, \ldots, \mathbf{d}_n$ from Gaussian distribution
- Estimate parameters

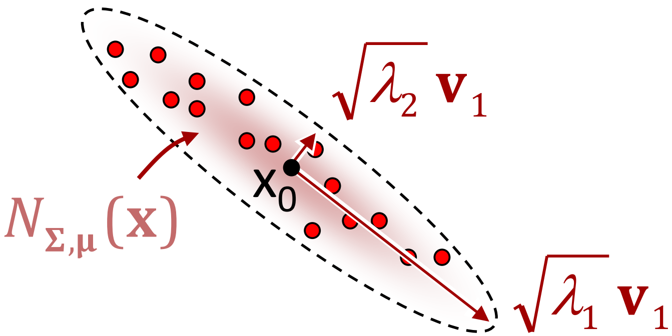
Maximum Likelihood Estimation

$$\boldsymbol{\mu}_{ml} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i \qquad \boldsymbol{\Sigma}_{ml} = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{d}_i - \boldsymbol{\mu})(\mathbf{d}_i - \boldsymbol{\mu})^{\mathrm{T}}$$

Principle Component Analysis (PCA)

- Fit Gaussian (ML-estimate)
- Compute main axes

PCA: Statistical Interpretation



$$\boldsymbol{\mu}_{ml} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i$$

$$\boldsymbol{\Sigma}_{ml} = \frac{1}{n-1}\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{d}_i - \boldsymbol{\mu})(\mathbf{d}_i - \boldsymbol{\mu})^{\mathrm{T}}$$

$$\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) := \left(\frac{1}{(2\pi)^{-\frac{d}{2}}\det(\boldsymbol{\Sigma})^{-\frac{1}{2}}}\right) e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

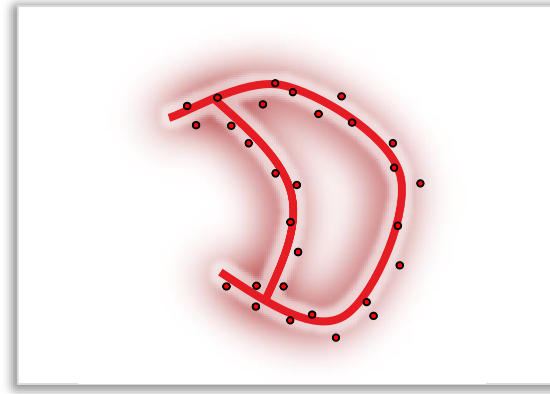
Statistical Interpretation of PCA

- Fitting a Gaussian, then compute its main axes
- $\frac{1}{n-1}\mathbf{S}$ is the covariance matrix of data $D = \{\mathbf{d}_1, \ldots, \mathbf{d}_n\}$
- $\boldsymbol{\mu}$ is the mean

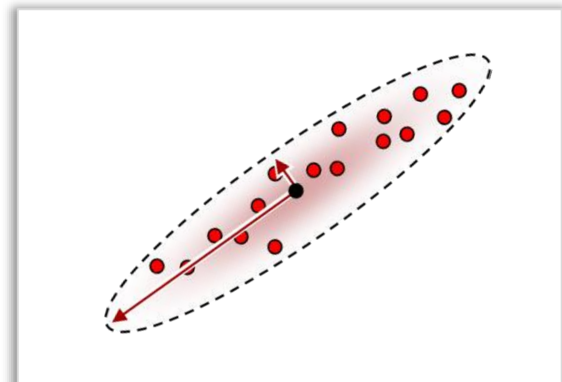
Three Different Stories!

PCA, told in 3 different ways…

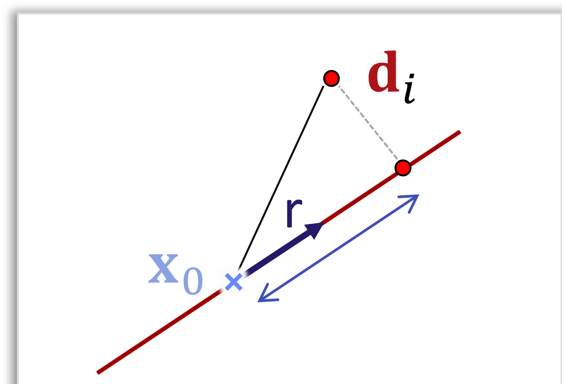
- **Total least-squares:**
  Euclidean error



- **Multivariate Gaussians:**
  Main Axes of Variance



- **Dimensionality Reduction:**
  Optimal projection to subspace



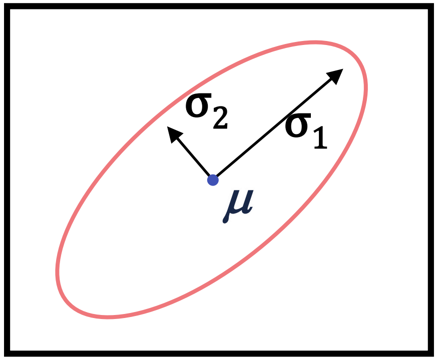
Quadratic Forms, Quadrics & Ellipses

The Shape of Gaussians

Probability Density

$$\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) := \left(\frac{1}{(2\pi)^{-\frac{d}{2}}\det(\boldsymbol{\Sigma})^{-\frac{1}{2}}}\right) e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

normalization



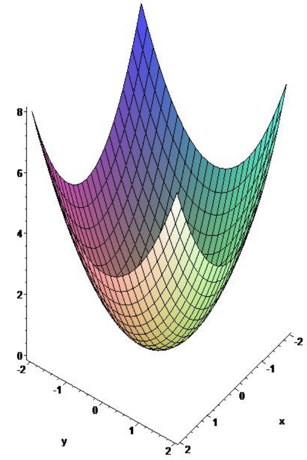
Neg-Log Density:

- $\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) + const$

normalization

Geometry

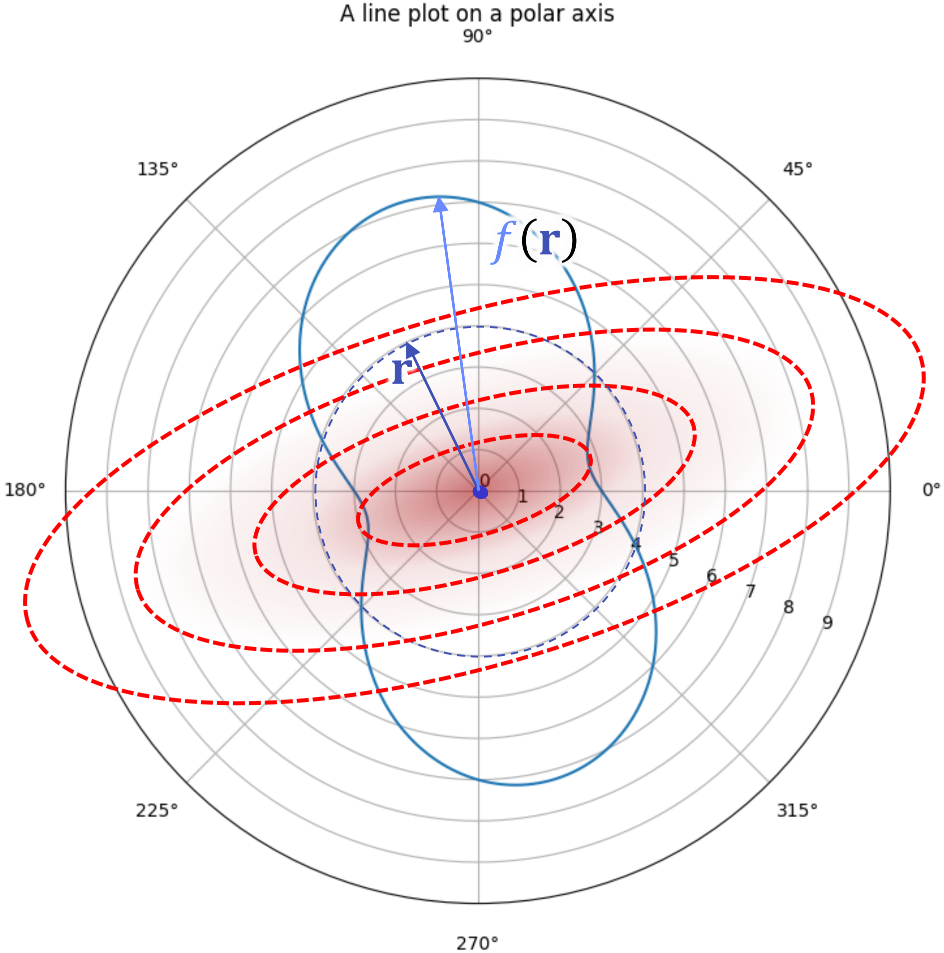
- Iso-probability profiles are ellipsoids
- Eigenvectors of Σ are main axes



Output by Direction

Example

$$f(\mathbf{r}) = \mathbf{r}^T \begin{pmatrix} 7 & 1 \\ 1 & 3 \end{pmatrix} \mathbf{r}$$

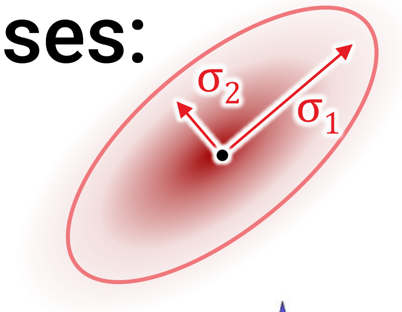


$$\|\mathbf{r}\| = 1$$

$$\approx \mathbf{r}^T \begin{pmatrix} 0.97 & 0.23 \\ -0.23 & 0.97 \end{pmatrix} \begin{pmatrix} 7.24 & 0 \\ 0 & 2.76 \end{pmatrix} \begin{pmatrix} 0.97 & -0.23 \\ 0.23 & 0.97 \end{pmatrix} \mathbf{r}$$

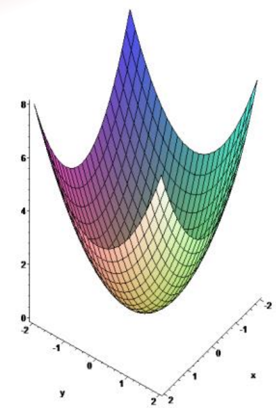
Reciprocals are Confusing

Level-sets of quadratic forms are ellipses:



- $\lambda^2 x^2 - c = 0$ … $\mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x} - c = 0$
- Larger eigenvalues → shorter radius

"Size of the Gaussian by EV"



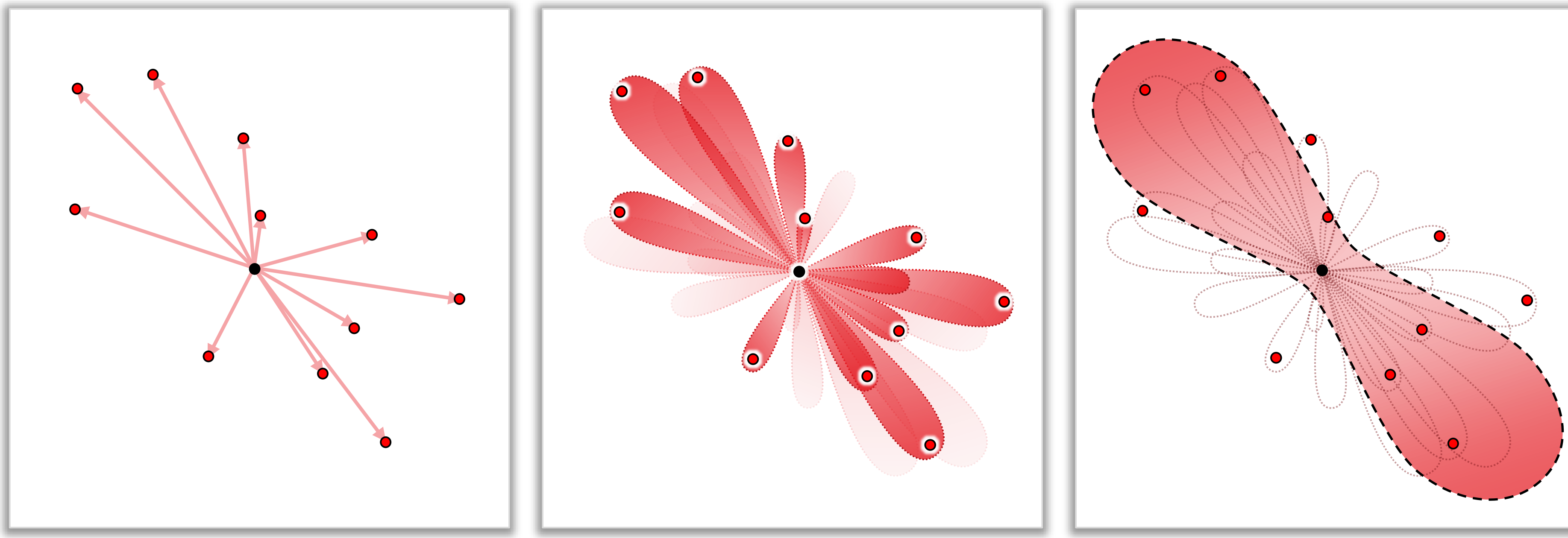
- $x \mapsto \exp(-\lambda^2 x^2)$ … $\mathbf{x} \mapsto \exp\left(-\mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x}\right)$
- Larger eigenvalues → smaller Gaussian

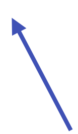
"Size of the Gaussian by variance/std-dev"

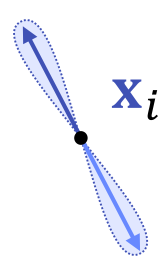
- $x \mapsto \exp(-x^2/\sigma^2)$ … $\mathbf{x} \mapsto \exp\left(-\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)$
- Larger variance → larger Gaussian

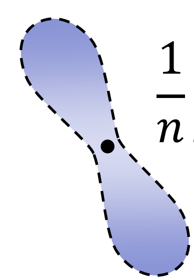
One More Story…

Geometric Interpretation of PCA: Averaging Σ



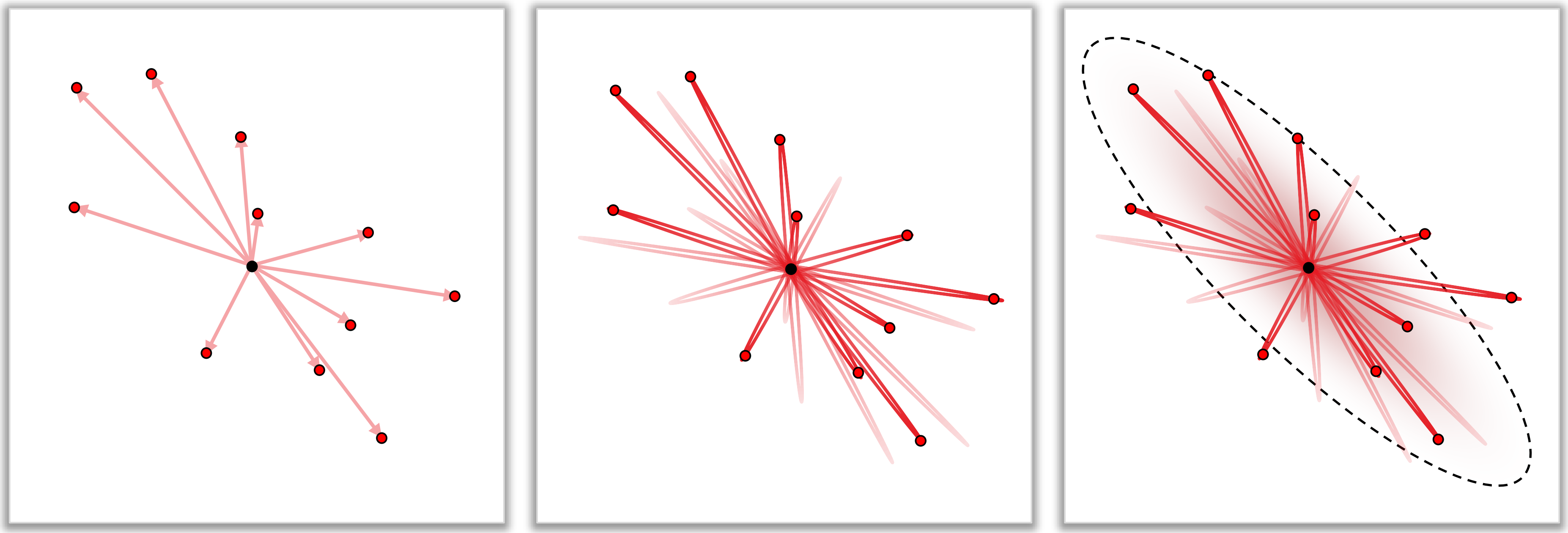

$\mathbf{x}_i \in \mathbb{R}^d$


$\mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \in \mathbb{R}^{d \times d}$


$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \in \mathbb{R}^{d \times d}$

Covariance Matrices

Level Sets of Σ^{-1}

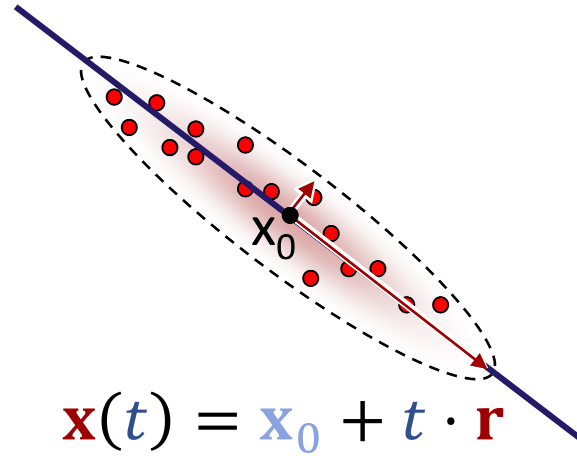


Applications

Applications

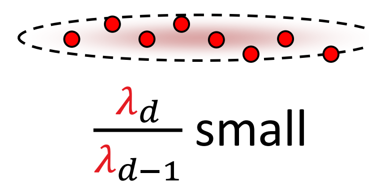
Fitting a line to a point cloud

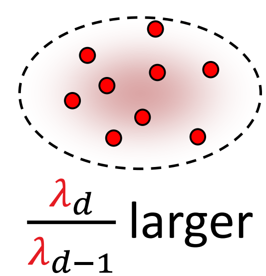
- Sample mean and direction of maximum eigenvalue



Plane Fitting in $\mathbb{R}^3$:

- Smallest eigenvalue: *normal* direction
- Aspect ratio λ_3/λ_2 is a measure of "flatness" (quality of fit)





Applications

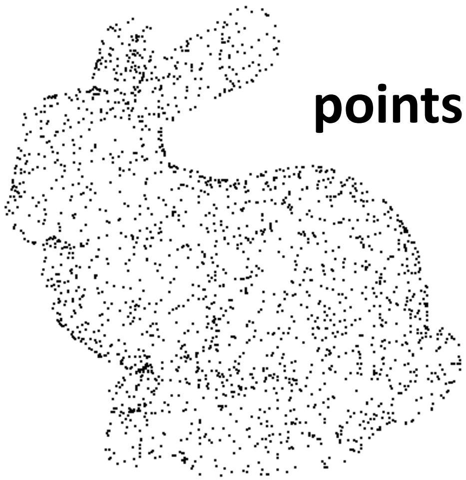
Normal estimation in point clouds

- Given a set of points $\mathbf{d}_1, \ldots, \mathbf{d}_n \in \mathbb{R}^3$
  - Forms a smooth surface
- Estimate
  - Surface normals
  - Sampling spacing

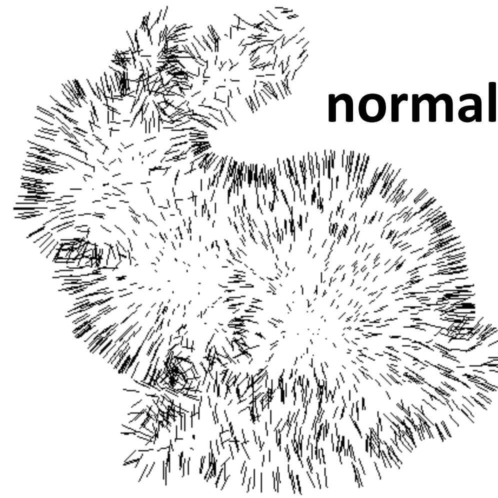
Algorithm:

- For each point: compute the k nearest neighbors ($k \approx 20$)
- Compute a PCA (average, main axes) of these points
  - Eigenvector with smallest eigenvalue $\rightarrow$ normal direction
  - The other two eigenvectors $\rightarrow$ tangent vectors
  - Tangent eigenvalues give sample spacing estimate

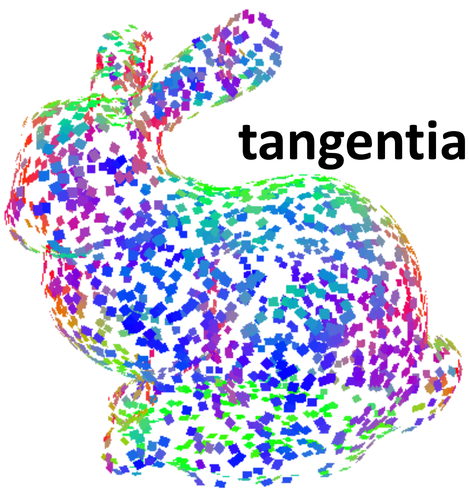
Example



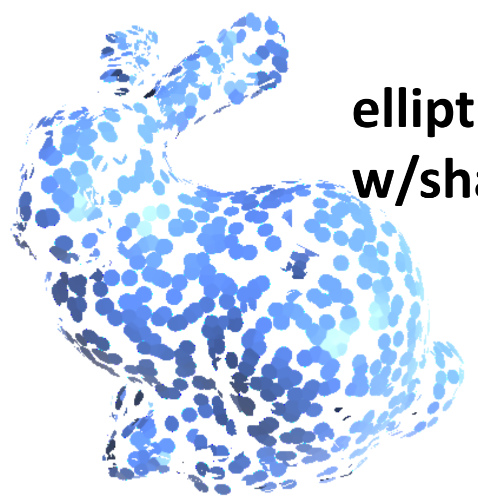
points



normals

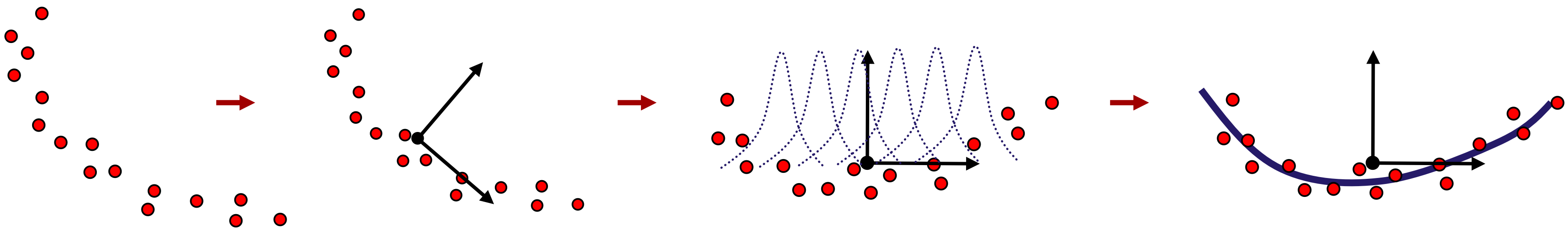


tangential frames



elliptic splats w/shading

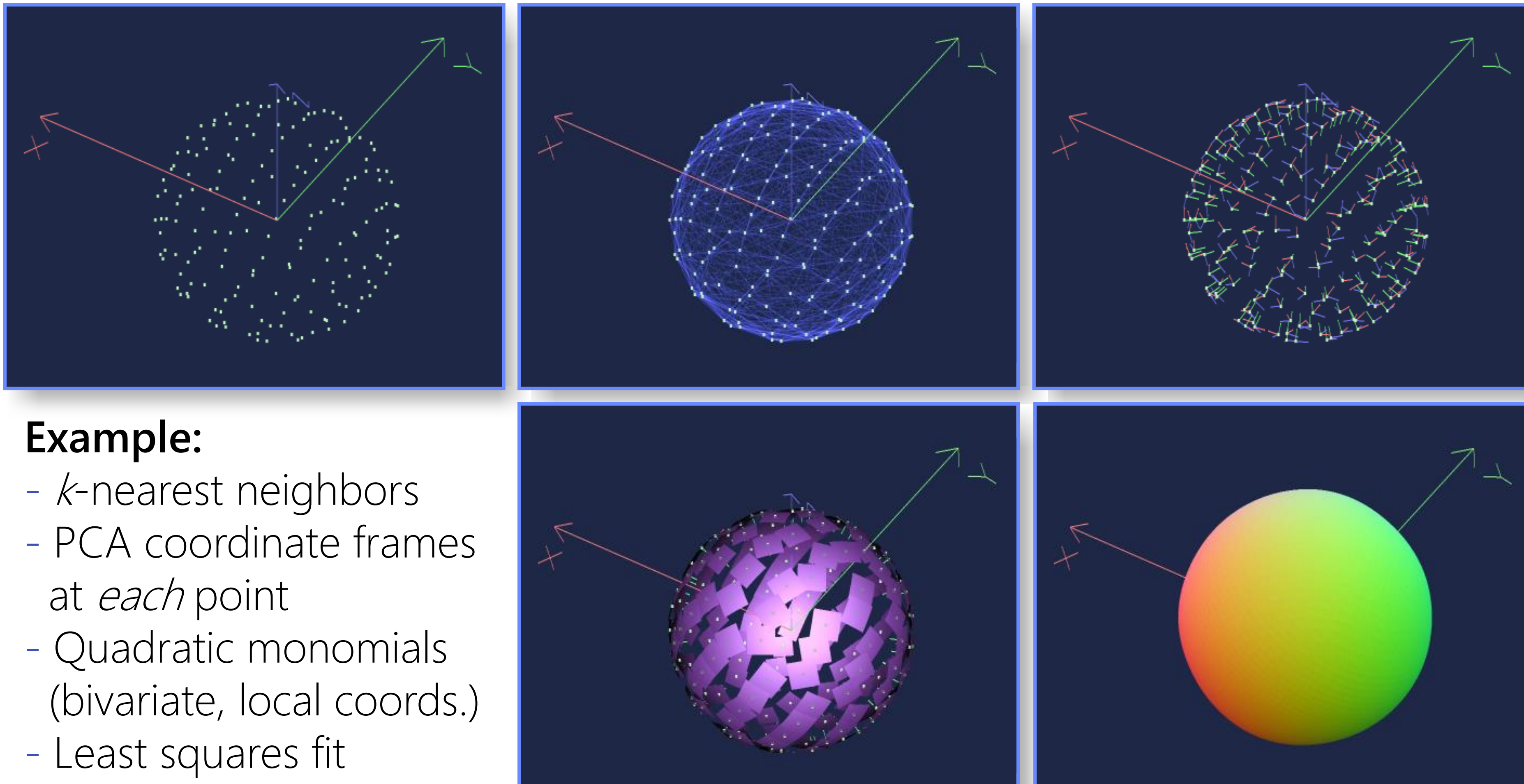
Applications



Another Application: Coordinate frame estimation

- More general total least-squares (non-linear) is difficult
- However: *We can tweak it*
  - Compute coordinate frames using PCA
  - Smallest eigenvalue = normal direction
  - Form height field in normal direction
  - Then: use ordinary least-squares in this coordinate system

Example



Example:
- *k*-nearest neighbors
- PCA coordinate frames at *each* point
- Quadratic monomials (bivariate, local coords.)
- Least squares fit

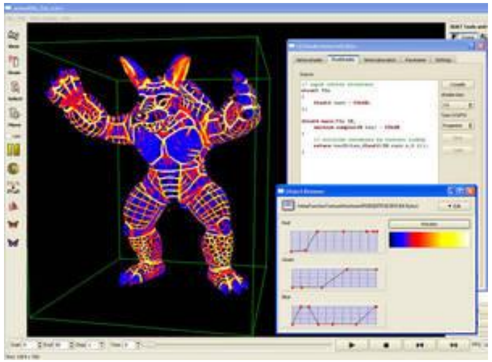
Curvature Estimation

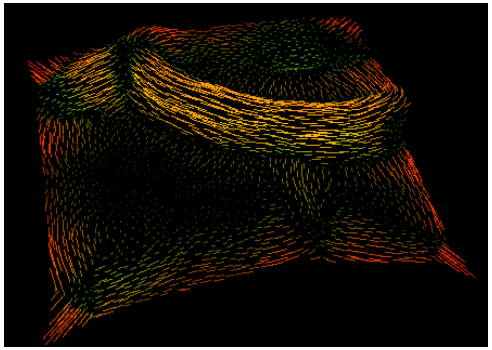
Estimating curvature in point clouds:

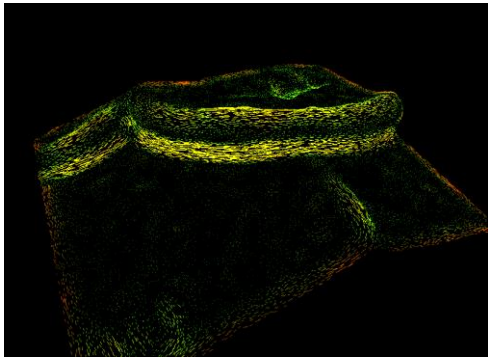
- Quadratic fitting (as in the example)
- Eigenanalysis of the quadratic terms
  - Hessian matrix from fitted polynomial

$$2\lambda_{uv}uv + \lambda_{uu}u^2 + \lambda_{vv}v^2 \rightarrow \begin{pmatrix} \lambda_{uu} & \lambda_{uv} \\ \lambda_{uv} & \lambda_{vv} \end{pmatrix}$$

  - Eigenvalues correspond to curvature
  - More on this in another lecture...

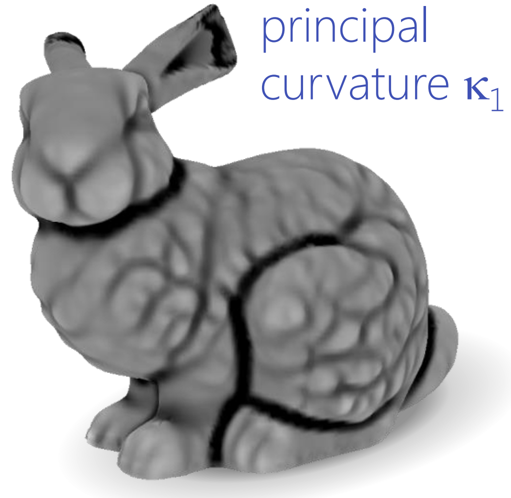




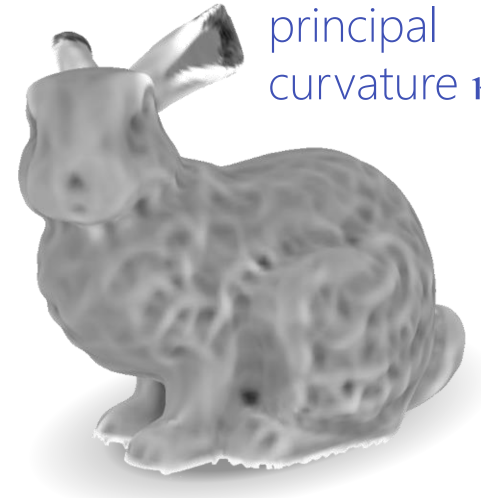


Bunny Curvature

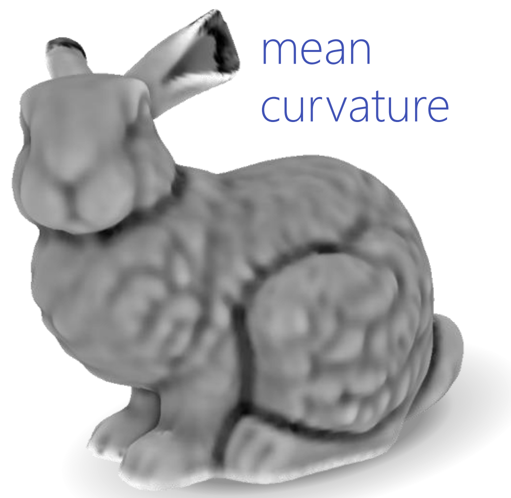
Stanford Bunny
(dense point cloud)



principal
curvature κ_1



principal
curvature κ_2



mean
curvature

Gaussian
curvature

[courtesy of Martin Bokeloh]

PCA on Face Spaces

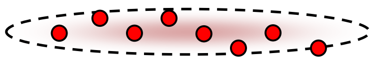
Volker Blanz & Thomas Vetter:
A Morphabable Model for the Synthesis of 3D Faces,
ACM Siggraph 1999.

http://mi.informatik.uni-siegen.de/sites/projects_01.php

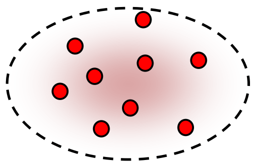
Outlier Removal for Smooth Surfaces

Plane Fitting in $\mathbb{R}^3$:

- Aspect ratio λ_3/λ_2 is
  a measure of "flatness" (quality of fit)
- Bad aspect ratio: not a surface point

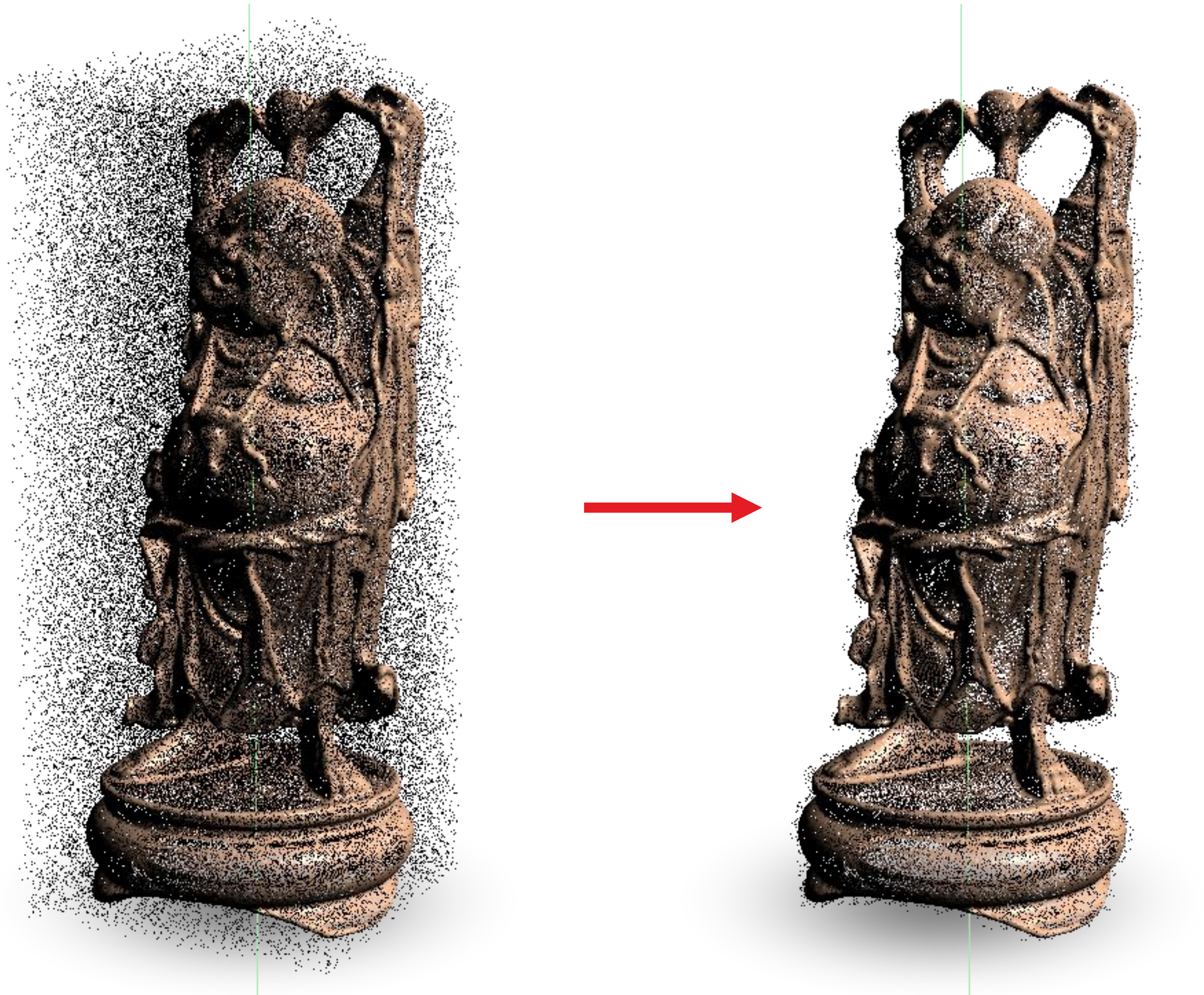


$\frac{\lambda_d}{\lambda_{d-1}}$ small

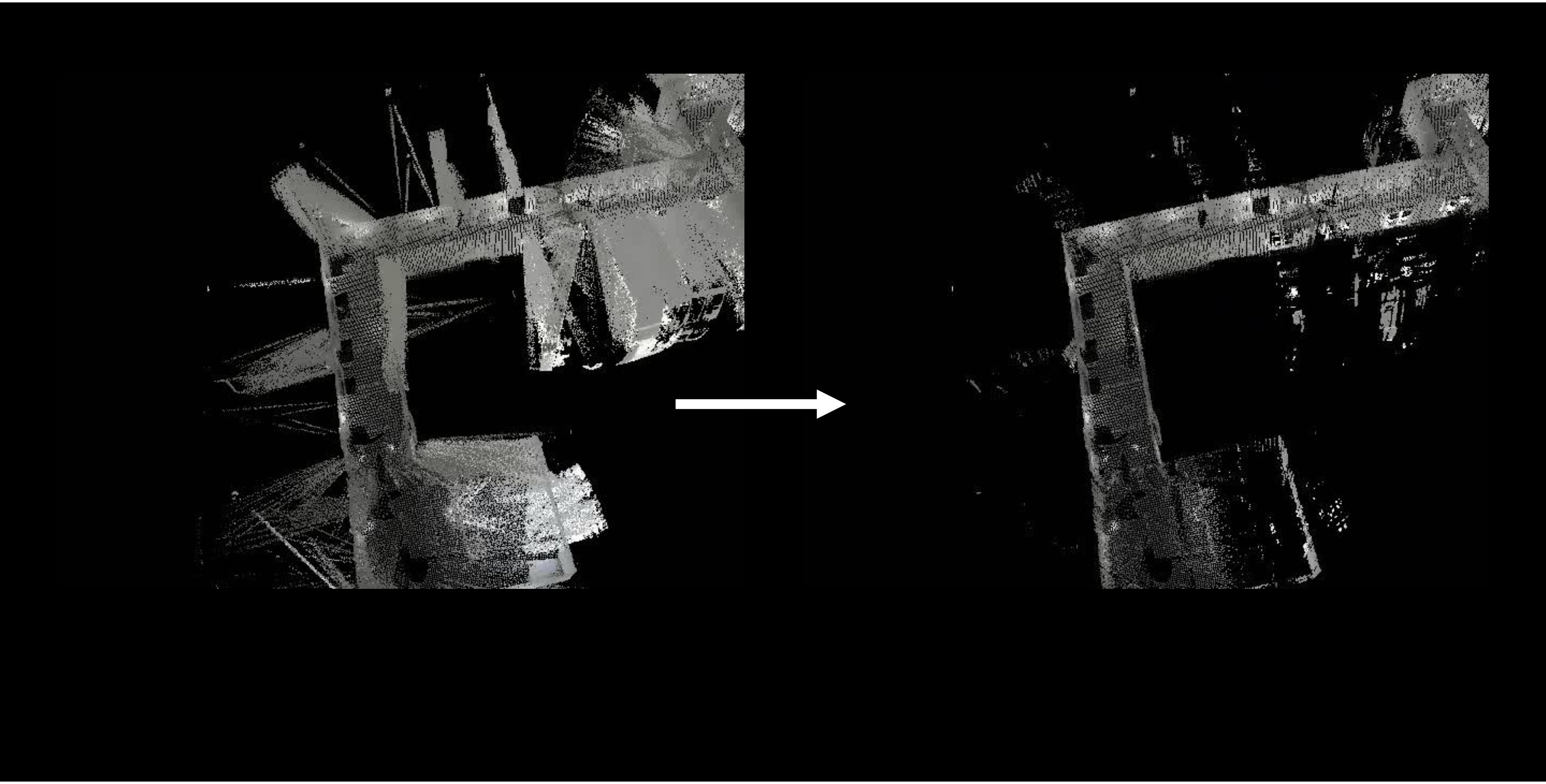


$\frac{\lambda_d}{\lambda_{d-1}}$ larger

Automatic Outlier Removal



Automatic Outlier Removal



Automatic Outlier Removal